Network Friendly P2P Streaming: The NAPA-WINE Architecture

Network Friendly P2P Streaming: The NAPA-WINE Architecture

E.Leonardi[1], M.Mellia[1], C.Kiraly[2], R. Lo Cigno[2], S.Niccolini[3], J.Seedorf[3]

[1]Politecnico di Torino, Italy; [2]University of Trento, Italy; [3]NEC Research Laboratories, Heidelberg, Germany

E-mail: [1]{emilio.leonardi,marco.mellia}@polito.it, [2]{kiraly,locigno}@disi.unitn.it, [3]{seedorf,niccolini}@nw.neclab.eu

Abstract: **Streaming video on P2P overlays puts extremely high demands and stress on the underlying network, especially in case of TV and live streaming. The NAPA-WINE consortium has devised an overall architecture for live video streaming that supports the needs of the users and content providers, while being respective of network-level needs, as reducing inter-AS traffic using ALTO-like services. Prototype applications following the proposed architecture and based on software libraries developed under GPL and LGPL licences by the consortium have already been implemented and are running both in partners premises and in demos showing the viability and performance of the solution.**

Keywords: P2P TV, Live Streaming, Network Friendly, ALTO

1 INTRODUCTION

In recent years, Peer-to-peer (P2P) technology became increasingly popular for video streaming applications, including TV services (P2P-TV). Examples of commercial P2P-TV include SopCast, TVAnts, PPLive, UUSee, and TVUplayer. The interest of the research community, content providers and network operators is also on the rise, though probably due to different reasons. Content providers see a novel opportunity to reach users, but at the same time they are concerned about the threats posed to their standard business models. Network operators are mainly worried by the stress posed by such a bandwidth-hungry and delay sensitive application on their infrastructure. The research community is stimulated by the opportunities offered by live P2P distribution and broadcasting [1], looking both for novel technical solutions and innovative business models.

NAPA-WINE (Network Aware P2P-TV Application over WIse NEtworks) is a three years project (STREP) within the 7-th Research Framework of the European Commission whose goal is finding innovative solutions for P2P live streaming to meet opportunities envisaged by content providers while soothing worries of network operators. Cooperation between network providers and P2P applications has been already proposed (e.g. [2]) and large consortia as the P4P project are considering it, but these works mainly address file sharing P2P applications, while P2P-TV or live-streaming in general have been neglected, probably due to the inherent difficulties of addressing a system with such tight performance demands.

In a P2P-TV system, a source divides the video stream into chunks of data, which are exchanged among nodes to distribute them to all participating peers. Peers form an overlay topology at the application layer, where neighbor peers are connected and exchange chunks using the underlying IP network. The IP and overlay layer are both "network" layers in that they both perform routing and forwarding of the data: packets in the IP layer and chunks (normally a sequence of packets) in the overlay.

In this context, the NAPA-WINE project proposes an innovative, network cooperative P2P architecture that explicitly targets the optimization of the quality perceived by the users while minimizing the impact on the underlying transport network. NAPA-WINE does not impose any structure on the overlay topology, which can be any type of generic mesh. The video distribution is chunk-based, but chunk construction is free enough to accommodate anything from a single video frame (even less if required) to large swaths of a video in case of nearly on-demand applications. The focus is on the design of a system empowering future P2P High Quality TV, a system where a source peer produces the video stream, chops it into chunks, and injects them in the overlay where peers cooperate to distribute them, without the need for the source to have enormous resources and bandwidth to support the service.

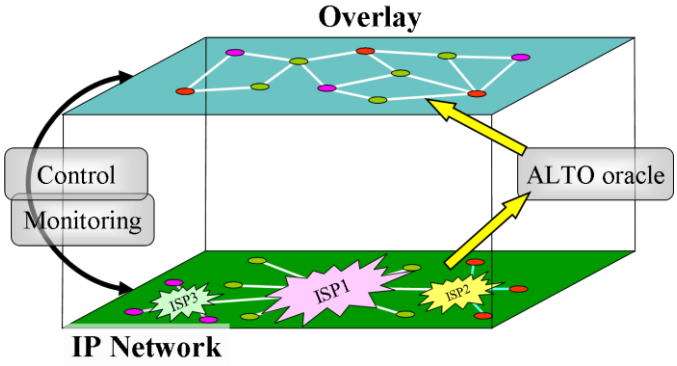

Figure 1: NAPA-WINE Vision - IP routers and overlay nodes (peers) cooperate through monitoring and control capabilities, and ALTO interaction to optimize performance and network usage

The architecture we envision is schematically represented in Fig. 1. The overlay and the IP network interact through monitoring and control capabilities with the aim of guaranteeing good quality to users and efficient use of resources to network operators. An additional element, not

Corresponding author: Renato Lo Cigno, University of Trento, Via Sommarive 14, Povo, 38123 Trento, Italy; Tel.+390461282026, e-mail:locigno@ disi.utnin.it

This work is supported by the European Commission through the NAPA-WINE Project (Network-Aware P2P-TV Application over Wise Network), ICT Call 1 FP7-ICT-2007-1, 1.5 Networked Media, grant No. 214412 – http://www.napa-wine.eu

mandatory, but definitely useful to optimize performance is the presence of an Application Layer Traffic Optimization (ALTO) oracle. As pursued by the IETF in the ALTO [3] working group with the contribution of NAPA-WINE partners, the network provider is given the capability to guide the P2P application, for example by explicitly publishing information about the status of its network, like link congestion or AS routing preferences.

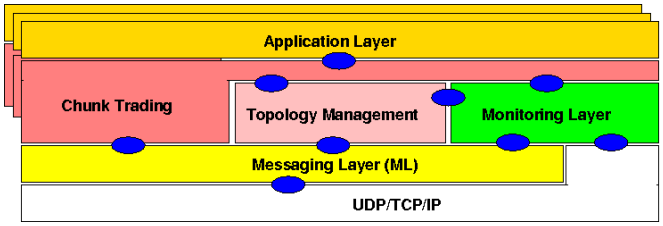

Figure 2: Logical architecture of a NAPA-WINE peer

The schematic representation of the NAPA-WINE protocol architecture is represented in Fig. 2, showing the functional blocks of a generic peer. An important element of the system is the built-in distributed monitoring tool that allows the application to continuously gather real time information both on network conditions and users' perceived performance. Information collected by the monitoring tool can be used to trigger reconfiguration of the overlay or to drive the scheduling of the chunk distribution protocol.

2 ARCHITECTURE OVERVIEW

The architecture depicted in Fig. 2 is based on four main building blocks, plus the external ALTO server that can support the topology management providing information that cannot be measured at the application level. In the following we briefly outline the role and key features of each building block.

2.1 User Layer

The User Layer is mainly responsible for of video encoding and its packaging into chunks (at the source) and de-chunkization and decoding at receivers. Standard encoding tools like ffmpeg can be used by the User Layer, whose goal is not implementing a proprietary video encoder, but supporting as many as possible standard formats (MPEG1/2/4, H.264, etc.). Depending on the type of video source this may include analog/digital conversion, encoding and any other video manipulation that the content provider whish to do, like advertising introduction and similar. A fundamental and distinctive feature of the NAPA-WINE User Layer is the flexible chucking supported. Most standard P2P-TV system tend to either map one frame in one chunk, or to blindly cut the stream into constant size chunks, oblivious of any media characteristics. Both solutions are sub-optimal: the first one may lead to very small chunks posing high demands on the chunk trading unless a structured overlay is built; the second one leads to severe degradation when one chunk is lost, since the frame boundaries are violated and normally players badly handle this situation. The chunkizer of the User Layer enables instead smart chunking techniques, like for instance collecting frames with different characteristics (e.g., I, P and B frames of MPEG flows) to be collected into separate chunks, that can be also treated with appropriate priorities by the chunk trader.

When the peer acts as receiver only, the user module reassembles the video stream from the series of received chunks, so that, after decoding, the video can be displayed.

Several instances of the User Layer can be combined within the same application allowing to watch a channel while recording another one, or having previews of different channels on the screen.

2.2 Peer Sampling and Topology Management

A P2P-TV client needs to communicate very efficiently with other peers to receive and redistribute the huge amount of information embedded in a video stream. Information must arrive in nearly real-time and with small delay variation.

The application goal is then to deliver all the video information to all peers in the system in the smallest possible amount of time. One of the key enabling factors is who are the peers to communicate with, i.e. the neighborhood selection.

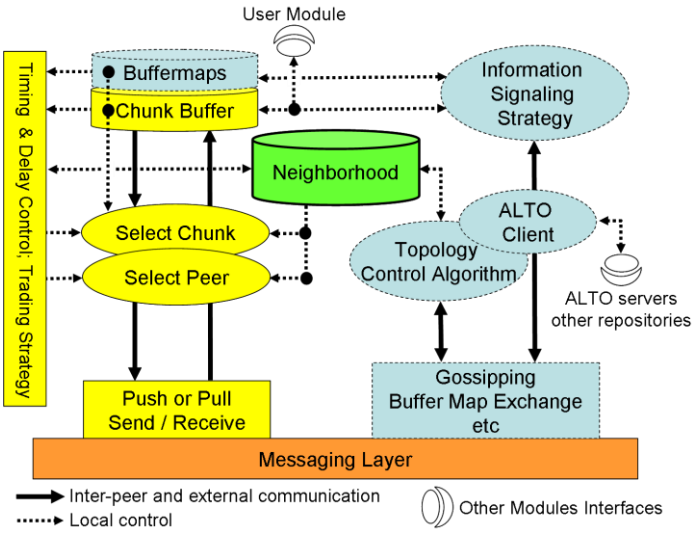

Figure 3: Detailed architecture of the topology management and the chunk trading logical layers

Fig. 3 depicts the logical relationship between the functions that compose the overlay management, and the functions that compose the chunk trading (see Sect. 2.3). Interfaces toward the other layers and external services like ALTO are also indicated.

2.2.1 Populating the Neighbourhood database

The Neighbourhood database (in green with thick lines in Fig. 3) is populated as soon as a peer participates in the distribution of a TV channel. For the sake of clarity we describe the architecture as if only one channel is present, but as shown in Fig. 2, multiple User Layers and multiple Overlays can be present to manage multiple channels. Referring to Fig. 3, the components in yellow (thin lines) are related to chunk scheduling, transmission and reception described later, while those in blue (dashed lines) refer to

topology management and signalling in general (exchange of buffer maps, i.e., the list of chunks available at each peer, availability to service chunks, etc.). The two functionalities interact through the Neighbourhood database as well as the chunk buffer (i.e., the structure where chunks are stored for trading and before playout), and the related buffer maps of neighbors.

The overlay network in P2P systems is the result of a distributed algorithm that builds and maintains the neighbourhood at each peer. When a peer joins the system, it selects an initial set of neighbours, then the set of neighbours of every node in the system is dynamically optimized over time.

The bootstrapping phase can be helped by the source or a web server where the user selects a channel, which can behave as a bit-torrent like tracker. The maintenance of the topology is based on a gossiping protocol like Newscast [4] or Cyclon [5] (both supported and implemented in the NAPA-WINE peers) that enables discovery of peers in the overlay. Once peers are discovered, the optimal topology management is obtained through an ad-hoc implementation of Tman [6], which has been tailored and optimized for live streaming and to interact with ALTO services when needed.

2.2.2 ALTO Support

The topology management in the NAPA-WINE architecture fully supports ALTO guidance through the integration of an ALTO client within the topology manager (see Fig. 3). Application Layer Traffic Layer Optimization [3, 7] is an innovative approach that enables network operators to save operational costs by reducing the amount of application layer backbone traffic. Essentially, ALTO is a dedicated service, operated by either ISPs, Content providers or by an independent provider, which provides useful network layer information to application layer clients about costs and resources. The IETF has formed a working group and is currently standardizing an ALTO protocol [7, 8]. Fig. 4 (reproduced from [3]) shows the overall idea behind such an ALTO service. Assume that Client 2 in the figure wants to connect to a stream. The gossiping protocol provides Client 2 several candidate peers which can offer the desired stream. In the figure, Client 2 can connect to Client 1 or Client 3. Client 2 queries an ALTO service for guidance on which Client to select. The ALTO service can answer queries based on information provided by the Client's ISP or the Content Provider. Queries can regard the topology, routing state, policies, or operational costs. In the example depicted it is likely that the ALTO service would suggest Client 3 because this Client is physically located in the same network as Client 2 and its choice will reduce inter-AS traffic.

2.3 Chunk Trading

Strictly related to topology management is the problem of chunk trading, which is the reason why the logical blocks have been depicted together in Fig. 3. The goal of chunk trading is receiving the stream smoothly (and with small delay) and to cooperate in the distribution procedure. We assume that peers are honest and cooperative, since the focus of NAPA-WINE is performance and system optimization, but we are well aware of the problems related to security, privacy, and cooperation in such a system [9].

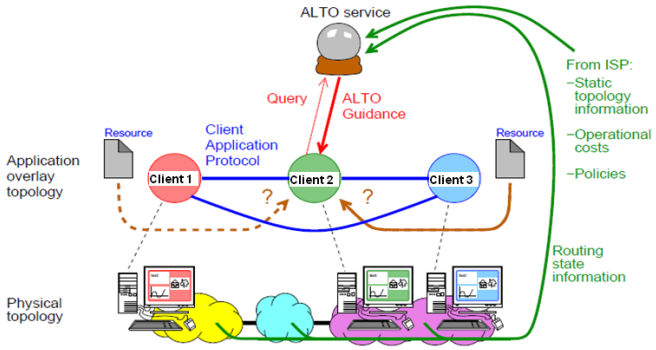

Figure 4: ALTO enabled peer selection

Chunks transferring is the operation that most affects performance and network friendliness. It includes protocol and algorithmic problems. First of all, peers need to exchange information about their current status to enable scheduling decisions. The information exchanged refers to the state of the peer with respect to the flow, i.e., a map of which chunks are needed by a peer to smoothly playout the stream. This task is carried out by the Information Signalling Strategy block in Fig.3. This block is in charge of i) sending buffer maps to other nodes with the proper timing, ii) receiving them from other nodes and merging the information in the local buffer map data base, iii) negotiating if this and other information should be spread by gossiping protocols or not, and to which depth it should spread in the topology.

Besides the buffer map exchange, the signalling includes Offer/Request/Select primitives used to trade chunks. These messages can be piggybacked on chunks for efficiency.

Another key protocol decision is about *Pushing* or *Pulling* information. A chunk is pushed when the peer owning the chunk decides to offer it to some other peer, while it is pulled when a peer needing the chunk requests it from another peer. From a theoretical point of view, as shown in [10], pushing is more effective.

Regardless of the protocol and the signalling strategy used, the core of a scheduler is the algorithm to choose the chunks to be exchanged and the peers to communicate with. Many different strategies have been studied, including both fundamental algorithms and their adaptation to heterogeneous scenarios, multiple sub-streams etc. [10,11,12,13]. The resulting algorithms are available in the prototype to experiment with.

2.4 Monitoring Layer

Beside the information provided by the ALTO Server, both the chunk scheduler and the overlay manager can exploit timely information about the quality of the connectivity between peers collected in real time by the monitoring modules. This includes, but is not limited to, the distance and

the available bandwidth between two peers, or the presence of Network Address Translation (NAT). The Monitoring Layer has two modes of operation: passive and active. Passive measurements are performed by observing the messages that are exchanged between peers. Active measurements, craft special probe messages which are sent to other peers at the discretion of the Monitoring Layer. The design and realization of the Monitoring Module is one of the innovative solutions of NAPA-WINE, as many measures like, available bandwidth in large scale systems, is far from trivial [14].

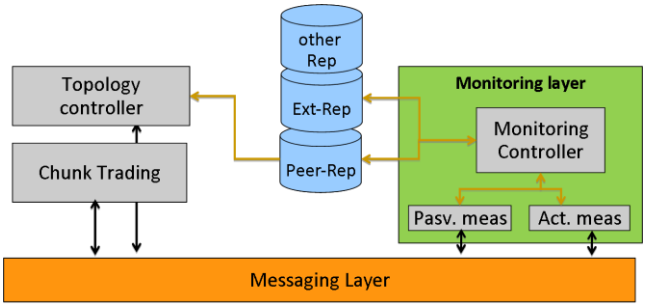

Figure 5: Detailed architecture of the monitoring modules.

2.4.1 Repositories

"Repositories" are the application level databases where information about each peer is shared with all other peers. Indeed, the information generated by the monitoring module is not only exploited by the local peer, but it is also made accessible to other peers that can benefit from it. The information acquired by each peer is stored locally and shared in the neighborhood database. It is also summarized and exported (published) to other peers or an external repository from which it can be retrieved by other peers, for instance at bootstrapping.

An ALTO service or oracle can be seen as a special case of external repository, which actually are a generalization of the ALTO concept. The local repositories can also be seen as an abstraction of the interaction of the topology management and chunk trading with the monitoring module.

2.5 Messaging Layer

The Messaging Layer (the bottom box in Fig. 3) offers primitives to other modules for sending and receiving data to/from other peers. It provides an abstract interface to transport protocols (UDP/TCP) and the corresponding service access points offered by the operating system by extending their capabilities and providing an application level addressing, i.e., assigning a unique identifier to each peer. For example, it provides the ability to send a chunk to another peer, which has to be segmented and then reassembled to fit into UDP datagrams. The messaging layer also provides an interface to the monitoring module invoked for passive measurements: whenever a message is sent or received an indication will be given to the monitoring module, which can update its statistics.

The last important feature of the messaging layer is mechanisms for the traversal of NAT boxes. Network Address Translation allows attaching several Computers to the Internet using only one globally unique IP address. Therefore it is very popular with private customers, who are also the main target audience for P2P TV. However, the presence of a NAT device may prevent peers from establishing connections to other peers. Therefore, special NAT traversal functions are offered by the messaging layer.

3 IMPLEMENTATION STATUS

The NAPA-WINE project is implementing all the features and logical blocks described in previous sections. The topology management and the chunk trading are implemented offering different instances and algorithms for experimental purposes. This has been obtained by designing a development toolkit and a set of libraries named GRAPES (Generic Resource Aware P2P Environment for Streaming) [15]. GRAPES is entirely written in C so that reuse and linking with any language is easy. It provides a set of building blocks that researchers can use, combine, and modify to test and compare algorithms and architectures, hopefully fostering the development of new ideas an applications. A first release of GRAPES is available at http://imedia.disi.unitn.it/P2PStreamers/grapes.html.

Other parts, like the Monitoring Layer, are released as stand alone libraries that can be reused also for non-streaming P2P applications.

Several streamers have already been implemented, exploring different algorithms and techniques. Fig. 6 presents a screenshot from one of these streamers, showing the streamed video together with the topology of the overlay, which is obtained applying ALTO guidance to obtain locality. On the left hand side of the figure a monitoring of the local chunk buffer evolution is displayed: during experimental runs, it helps understanding the evolution of the swarm and correlate objective application level metrics (chunks are missing!), with the Quality of Experience (QoE) that a user perceives: degraded video quality, bad audio or motion freezing due to frame synchronization errors.

Strictly related to the QoE evaluation, a set of tools, named PSNR Tools (available at http://imedia.disi.unitn.it/QoE/) have been developed to evaluate the objective quality of video in terms of PSNR (Peak Signal to Noise Ratio) and other similar measures reconstructing the video stream starting from a trace of received chunks and comparing it to the original non-encoded video. These tools can be used both in conjunction with the actual applications (provided that the original video stream is somehow available, e.g., by sending the trace of received/missed chunks back at the source) or coupled with the simulators developed for early algorithmic testing in NAPA-WINE, as done in [17].

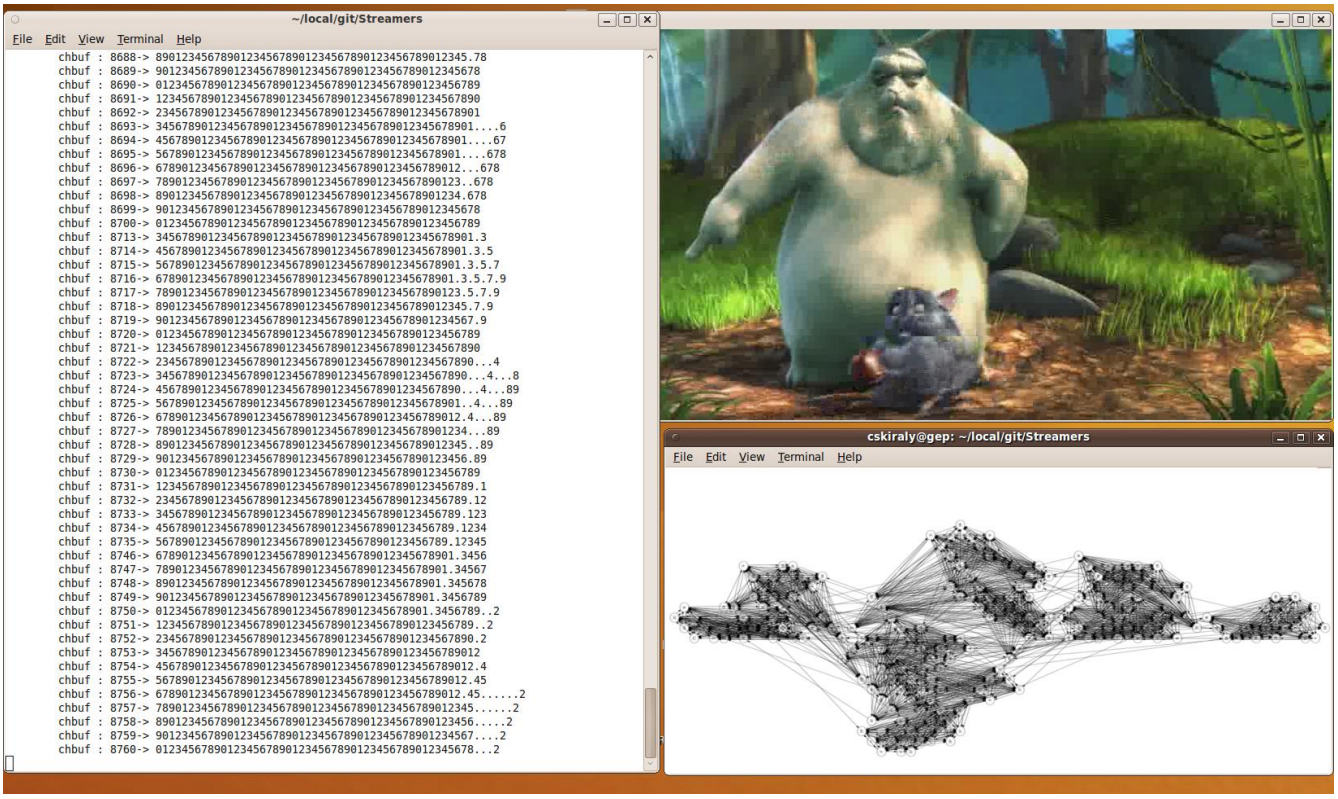

Figure 6: A Screenshot from a running peer showing the video, the topology of a the swarm exploiting locality from ALTO guidance (low left box), and the status of the local chunk buffer evolving in time

4 CONCLUSION AND DISCUSSION

Video Streaming applications exploiting the P2P communication paradigm are a commercial reality, but their overall architecture is still biased by file-sharing applications and they operate without any coordination with the IP network, often resulting in poor, even wasteful resource usage. This will prevent them to support High Quality TV in the future, or to make the transition to High Definition, which will require several Mbit/s per peer.

This paper has discussed the NAPA-WINE architecture for a P2P-TV system, which is under development in the project with the goal of efficiency and cooperation between the application and both the network operators and the content providers. Prototypes of the peers and system are already running on the Internet and are demonstrated at major venues [18,19]. The key features are the presence of continuous monitoring and control of the overlay, obtained also with ALTO support, whose oracle enables the exchange of information between the network and the application.

The overlay topology management and the chunk scheduling of information have been identified as important features for the application to be network-friendly. The first function enables building efficient and rational overlay topologies that are correctly mapped on top of the transport network structure (e.g., considering minimal number of hops between neighbours, locality w.r.t. Autonomous Systems, etc.). The second function guarantees that the network capacity is exploited without waste (e.g., by minimizing retransmissions and pursuing an efficient distribution of chunks, etc.).

The software and prototypes implemented in the NAPA-WINE project are made available as software libraries under GPL or LGPL license: they are evolving and are freely downloadable from the project or partners web sites.

Acknowledgments

We are deeply in debt with all the people and partners involved in NAPA-WINE and contributing to the project success with their work and research.

References

[1] L.Jiangchuan, S.G.Rao, L. Bo, H.Zhang, "Opportunities and Challenges of Peer-to-Peer Internet Video Broadcast." *Proceedings of the IEEE*, Vol.96, no.1, pp.11-24, Jan. 2008.

[2] V. Aggarwal, A. Feldmann, C. Scheideler, "Can ISPS and P2P users cooperate for improved performance?" *SIGCOMM Comput. Commun.Rev*. Vol.37, N.3, pp.29-40, Jul. 2007.

[3] J.Seedorf, S.Kiesel, M.Stiemerling, "Traffic Localization for P2P Applications: The ALTO Approach." *IEEE P2P 2009*, Seattle, WA, Sept. 2009.

[4] M. Jelasity, R. Guerraoui, A.-M. Kermarrec, M. van Steen, "The peer sampling service: Experimental evaluation of unstructured gossip-based implementations," *Middleware 2004*, LNCS 3231, Springer-Verlag.

[5] S. Voulgaris, D. Gavidia, M. Van Steen, "CYCLON: Inexpensive membership management for unstructured P2P overlays," *Journal of Network and Systems Management*, 13(2):197–217, 2005.

[6] M. Jelasity, A. Montresor, O. Babaoglu, "T-man: Gossip-based fast overlay topology construction," *Comput. Netw.*, 53(13):2321–2339, 2009.

[7] J. Seedorf and E. Burger, “Application-Layer Traffic Optimization (ALTO) Problem Statement,” RFC 5693, Oct. 2009
[8] R. Alimi, R. Penno, Y. Yang: “ALTO Protocol”, Internet Draft (work in progress), IETF, July 2010, http://tools.ietf.org/html/draft-ietf-alto-protocol.
[9] G. Gheorghe, R. Lo Cigno, A. Montresor, “Security and privacy issues in P2P streaming systems: A survey,” *Peer-to-Peer Networking and Applications* (on-line 23 April 2010), Springer
[10] L.Abeni, C.Kiraly, R. Lo Cigno, “On the Optimal Scheduling of Streaming Applications in Unstructured Meshes,” *IFIP Networking 2009*, Aachen, DE, May 11–15, 2009.
[11] A. Couto da Silva, E. Leonardi, M. Mellia, M. Meo, “A Bandwidth-Aware Scheduling Strategy for P2P-TV Systems”, *IEEE P2P 2008*, Aachen, DE, Sept. 2008.
[12] L. Abeni, C. Kiraly, and R. Lo Cigno, “Scheduling P2P Multimedia Streams: Can We Achieve Performance and Robustness?”, *IEEE IMSAA-09*, Bangalore, India, Dec. 9-11, 2009.
[13] L.Abeni, C.Kiraly, R. Lo Cigno, “Robust Scheduling of Video Streams in Network-Aware P2P Applications,” *IEEE ICC’09*, Cape Town, ZA, May 23–27, 2010
[14] D. Croce, M. Mellia, and E. Leonardi, “The Quest for Bandwidth Estimation Techniques for large-scale Distributed Systems,” *ACM Hotmetrics workshop, Sigmetrics 2009*, Seattle, CA, June 15-19, 2009.
[15] L.Abeni; C. Kiraly; A. Russo; M. Biazzini; R. Lo Cigno, “Design and Implementation of a Generic Library for P2P Streaming," Advanced video streaming techniques for peer-to-peer networks and social networking workshop, ACM Multimedia 2010, Oct. 25-29 Firenze, Italy
[16] A. Carta, M. Mellia, M. Meo, S. Traverso, “Efficient Uplink Bandwidth Utilization in P2P-TV Streaming Systems,” *IEEE GLOBECOM 2010*, Miami, FL, US, Dec. 7-9 2010
[17] L.Abeni, C.Kiraly, R. Lo Cigno, “"Deadline-based Differentiation in P2P Streaming,” *IEEE GLOBECOM 2010*, Miami, FL, US, Dec. 7-9 2010
[18] J. Seedorf, S. Niccolini, R. Lo Cigno, C. Kiraly , “Prototypical Implementation of ALTO Client and ALTO Server and Integration into a P2P Live Streaming Software”, Demonstration, *IPTComm 2010*, Aug. 3-4, Munich, DE
[19] L. Abeni, A. Bakay, M. Biazzini, R. Birke, E. Leonardi, R. Lo Cigno, C. Kiraly, M. Mellia, S. Niccolini, J. Seedorf, T. Szemethy, G. Tropea, “Network Friendly P2P-TV: The Napa-Wine Approach,” Live Demonstration and Extended Abstract, *IEEE P2P 2010*, Aug. 25-27, 2010, Delft, NL